

The spatial correspondence and genetic influence of interhemispheric connectivity with white matter microstructure

Jeroen Mollink^{1,2*}, Stephen M. Smith¹, Lloyd T. Elliott^{3,4}, Michiel Kleinnijenhuis¹, Marlies Hiemstra², Fidel Alfaro-Almagro¹, Jonathan Marchini^{1,4,5}, Anne-Marie van Cappellen van Walsum², Saad Jbabdi^{1,6} and Karla L. Miller^{1,6}

Microscopic features (that is, microstructure) of axons affect neural circuit activity through characteristics such as conduction speed. To what extent axonal microstructure in white matter relates to functional connectivity (synchrony) between brain regions is largely unknown. Using MRI data in 11,354 subjects, we constructed multivariate models that predict functional connectivity of pairs of brain regions from the microstructural signature of white matter pathways that connect them. Microstructure-derived models provided predictions of functional connectivity that explained 3.5% of cross-subject variance on average (ranging from 1–13%, or $r = 0.1$ –0.36) and reached statistical significance in 90% of the brain regions considered. The microstructure–function relationships were associated with genetic variants, co-located with genes DAAM1 and LPAR1, that have previously been linked to neural development. Our results demonstrate that variation in white matter microstructure predicts a fraction of functional connectivity across individuals, and that this relationship is underpinned by genetic variability in certain brain areas.

Communication between brain regions is achieved by axons grouped in white matter pathways. Properties of these structural connections are highly relevant to brain function, often described as functional connectivity. However, it is not simply the presence of a connection, but also the microscopic tissue architecture (that is, microstructure), of white matter that influences brain function. For example, axonal diameter, myelination and length all affect the precise timing of neural signals, which is crucial to synchronizing network dynamics¹.

Much of our knowledge about structural connectivity in the brain comes from animals², human lesions³ and postmortem human dissections⁴. These approaches have relatively high biological specificity and interpretability, but are limited in their ability to characterize inter-individual differences. More recently, diffusion magnetic resonance imaging (dMRI) has emerged as a powerful *in vivo* tool for studying the brain's structural connections⁵. Although limited in spatial resolution⁶, dMRI has the unique ability to estimate the trajectories of white matter bundles (that is, tractography) as well as certain microstructural properties of these bundles, through models linking the within-voxel dMRI signal to tissue architecture. An important benefit of dMRI is that it enables us to characterize inter-individual differences, even in large cohorts (for example, UK Biobank⁷). dMRI thus has the potential to relate individual variations in white matter microstructure to differences in brain function, which can also be characterized with MRI.

Diffusion and functional MRI have been used to investigate structure–function relationships, relating the anatomy of a white matter tract to the functional coupling between the regions it

connects^{8–11}. Importantly, these studies relate the macroscopic organization of the network to brain function but did not aim to establish whether the microstructural properties of a white matter tract relate to the functional communication it establishes between brain areas.

Several studies have demonstrated the potential for dMRI to establish relationships between microstructure and function. For instance, the commonly used metric fractional anisotropy (FA) is a measure of diffusion directionality that is biologically non-specific, being sensitive to many properties including axon density, size and myelination¹². Mean FA in a given white matter tract has been demonstrated to correlate with strength of functional connectivity^{13,14}. However, these studies focused on the tract connecting a single pair of regions and summarize a tract's microstructure with a single quantity (for example, FA averaged over the entire tract).

In this work, we address whether functional connectivity between brain regions is mediated by the microstructure of white matter pathways that connect them. We hypothesize that a data-driven model based on dMRI metrics can predict cross-subject variation in functional connectivity and, more specifically, that this is a general principle that holds across many brain regions and the pathways connecting them. Unlike previous literature, we generated models that capture rich spatial representation of a tract's microstructural profile (that is, a microstructural signature). In addition to diffusion tensor-based metrics, we incorporated estimates from a more sophisticated biophysical model that aims to provide greater biological specificity¹⁵. We consider interhemispheric connectivity between pairs of homotopic regions (that is, the homologous region in the two cerebral hemispheres) that are connected by

¹Wellcome Centre for Integrative Neuroimaging, University of Oxford, Oxford, United Kingdom. ²Department of Anatomy, Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Center, Nijmegen, the Netherlands. ³Department of Statistics and Actuarial Science, Simon Fraser University, Burnaby, British Columbia, Canada. ⁴Department of Statistics, University of Oxford, Oxford, United Kingdom. ⁵The Wellcome Centre for Human Genetics, University of Oxford, Oxford, United Kingdom. ⁶These authors contributed equally: Saad Jbabdi and Karla L. Miller.

*e-mail: jeroen.mollink@radboudumc.nl

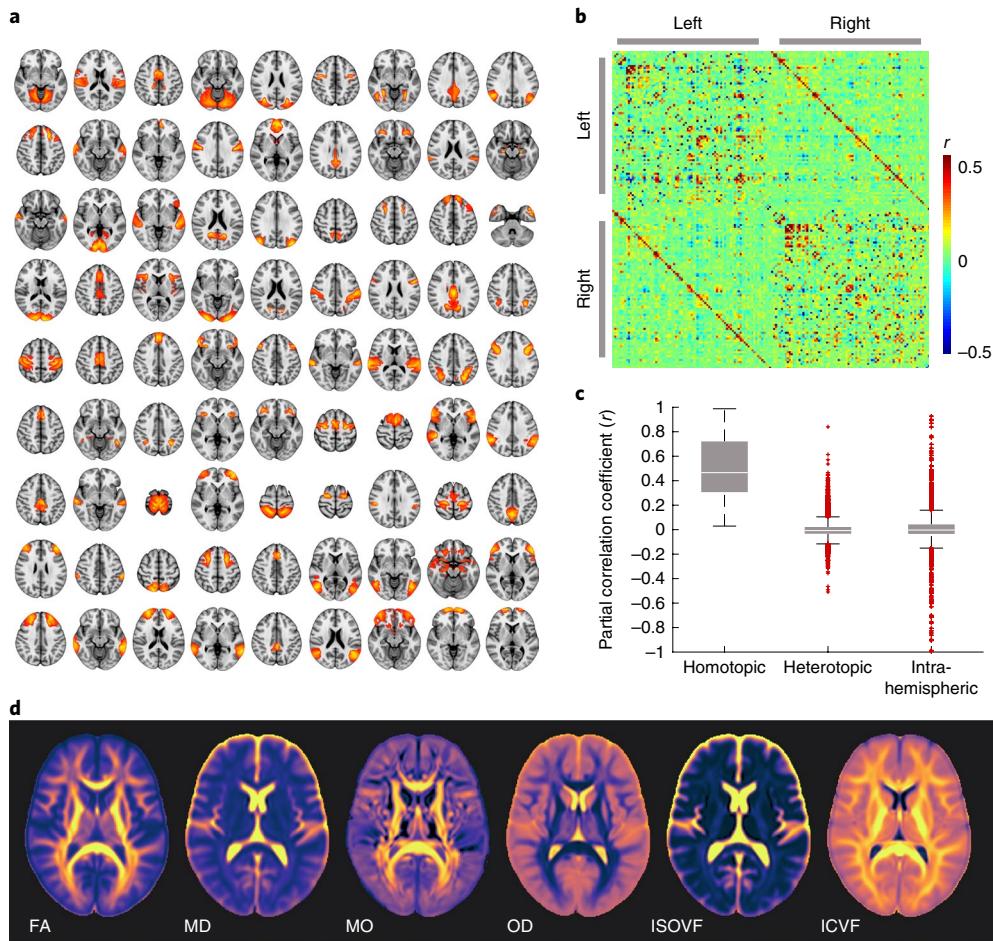


Fig. 1 | Definition of homotopic brain regions and dMRI-derived microstructural maps. **a**, Functional nodes were defined by applying ICA to the resting-state fMRI data, splitting between the hemispheres and isolating contiguous regions. These were then matched between hemispheres into 81 homotopic pairs, most automatically identified from the same independent component except for 10 manually identified from different components. **b**, Connectivity between homotopic pairs was estimated by partial correlation of the average time-series of each node, as shown in the connectome (matrix entries are sorted first by hemisphere and then by node number). **c**, Strength (partial correlation) of different functional connections in the brain, sorted by type. The center line depicts the median correlation coefficient for a specific type of connection; box limits, the 25th and 75th percentiles of the correlation coefficients; the whiskers extend to the most extreme data points excluding outliers (marked with a + symbol). Group-average estimates from $n=11,354$ subjects for $n=3,240$ connections evaluated. **d**, Group-averaged microstructure maps derived from the dMRI data. See text for explanation of abbreviations in **d**.

commissural white matter axons that run through the corpus callosum. We built a set of regression models to relate the tract's microstructural profile to functional connectivity for a large number of paired homotopic regions.

The models described above linking white matter microstructure to functional connectivity were trained ($n=7,481$) and replicated ($n=3,873$) on data from the UK Biobank⁷. We show that these models can predict up to 13% of the cross-subject variance in functional connectivity, and demonstrate that the microstructure–function link exists for a large number of brain regions and is highly reproducible. We additionally performed genome-wide association studies (GWASs) to identify single-nucleotide polymorphisms (SNPs) that are significantly associated with functionally relevant microstructure in the brain¹⁶. The identified SNPs are co-located with genes that have been reported to play an important role in axonal guidance and cortical development.

Results

In our primary analysis, we tested for microstructure–function relationships between homotopic brain regions and the callosal pathways connecting them, using dMRI and resting-state functional

MRI (fMRI) data from subjects in the UK Biobank project⁷. All subjects were selected based on usable resting-state fMRI and dMRI data, in addition to genetic inclusion criteria (see Methods). The activity of homotopic region pairs is often synchronized, with high functional connectivity^{17,18}. These pairs are primarily connected through the corpus callosum, the largest commissural pathway in the brain, which is well defined at typical imaging resolutions employed with dMRI.

Functional connectivity. We previously conducted a group-average decomposition of resting-state fMRI data using independent component analysis (ICA), which yielded 55 components corresponding to resting-state networks^{7,19}. For the work here, more finely grained functional ‘nodes’ were then generated from these components by first splitting each component into its constituent parts for right and left hemispheres, and further splitting if a component still contained non-contiguous brain areas. Homologous regions for the two hemispheres were then identified as nodes with strong similarity, producing 81 homotopic pairs (see Fig. 1a). Functional connectivity was estimated at the single-subject level by partial correlation of the average blood oxygen level-dependent signal

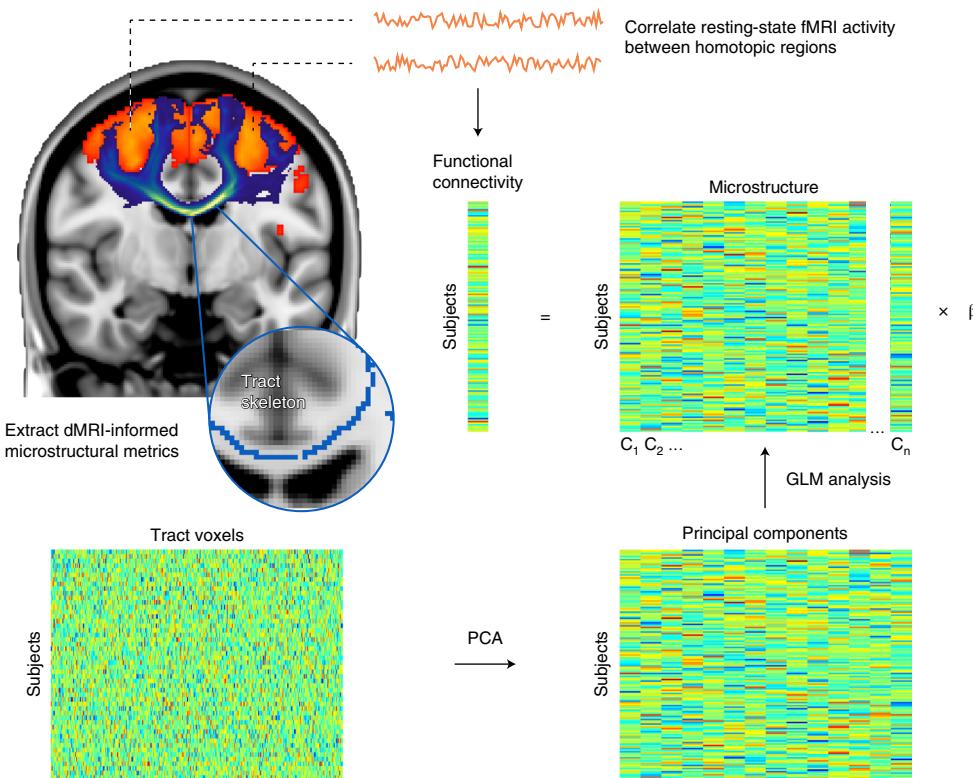


Fig. 2 | Prediction of functional homotopic connectivity from white matter microstructure. Between a pair of functionally defined homotopic areas (shown in orange in the brain), probabilistic tractography was performed to delineate the neuronal tract (shown in blue). The white matter skeleton voxels within a tract were stored as rows in a matrix for each subject. To extract the highest cross-subject variance among the tract-based spatial statistics (TBSS) voxels for a given microstructure metric, we performed a dimensionality reduction on this matrix using PCA. The top principal components ($n=30$) were fed into a linear regression model as explanatory variables for the functional connectivity between a homotopic pair.

time-series (equivalent to regressing out the time-series from all other regions before calculating pairwise correlations). This gives a connectivity matrix for each subject, which is summarized in Fig. 1b as the mean partial correlation across all subjects. Entries in this matrix are ordered first by hemisphere and then by region number, such that interhemispheric connections are given in the upper right and lower left quadrants. Homotopic connections, shown on the diagonals of these quadrants, were found to express on average the strongest connections in the brain, larger than intra-hemispheric or heterotopic interhemispheric connections (see Fig. 1c), in agreement with previous studies^{17,18}.

White matter microstructural signature. A range of microstructural features was derived from the dMRI data for the white matter pathway connecting each pair of homotopic gray matter regions. The diffusion tensor model describes the three-dimensional (3D) water displacement profile at each voxel using an ellipsoid²⁰. We extracted estimates of FA, mean diffusivity (MD) and anisotropy mode (MO)²¹ from this tensor fit. Neurite orientation dispersion and density imaging (NODDI)¹⁵ is a more biologically motivated model that aims to decompose the diffusion signal into an intracellular volume fraction (ICVF) and an isotropic volume fraction (ISOVF), the latter representing interstitial and cerebrospinal fluids. In addition, NODDI estimates an orientation dispersion (OD) index that quantifies the spread of fibers within the intracellular compartment. These dMRI-derived metrics represent an average across thousands of cellular components within each imaging voxel ($2 \times 2 \times 2 \text{ mm}^3$). Fig. 1d depicts a brain map of each microstructural metric averaged across all subjects. The white matter pathway that connects a given homotopic region pair was identified using probabilistic tractography²² performed on the dMRI data between the regions.

Predicting functional connectivity with microstructure. We performed a multiple regression analysis to test whether the microstructural features could predict cross-subject patterns of functional connectivity in the main cohort of 7,481 subjects. For a given homotopic pair of regions, the functional connectivity for all subjects was represented as a vector ($N_{\text{subjects}} \times 1$). To model the spatial patterns of white matter microstructure in a given tract, we begin by constructing a matrix that contains the dMRI-derived metric of interest for every subject (that is, a matrix $N_{\text{subjects}} \times N_{\text{voxels}}$). The included voxels are restricted to the center of the tract of interest using a standard ‘skeletonization’ procedure²³. Because these microstructure matrices are too large to robustly perform a direct regression ($N_{\text{voxels}} = 5,750 \pm 4,000$), we use principal component analysis (PCA) to reduce the matrix dimensionality. The top 30 principal components (see Supplementary Fig. 1) were extracted to serve as a set of regressors, resulting in an $N_{\text{subjects}} \times 30$ regression matrix (see Fig. 2 for an overview). Seven linear models were created for each homotopic pair: one for each of the dMRI-derived metrics (FA, MD, MO, OD, ISOVF, ICVF) and a multimodal approach combining all these microstructural metrics in a single matrix. For the multimodal analysis, the microstructural matrix for each metric was first normalized by its first singular value, and these normalized matrices were concatenated to form a single multimodal matrix (of size $N_{\text{subjects}} \times 6N_{\text{voxels}}$) that was again reduced to include only the top 30 principal components.

We first tested the hypothesis that dMRI-based microstructure can be used to predict cross-subject variation in functional connectivity consistently across many brain regions. We assessed the statistical significance of each model using permutation testing, performed independently across the homotopic pairs and models, and then corrected for multiple comparisons (see Methods). The

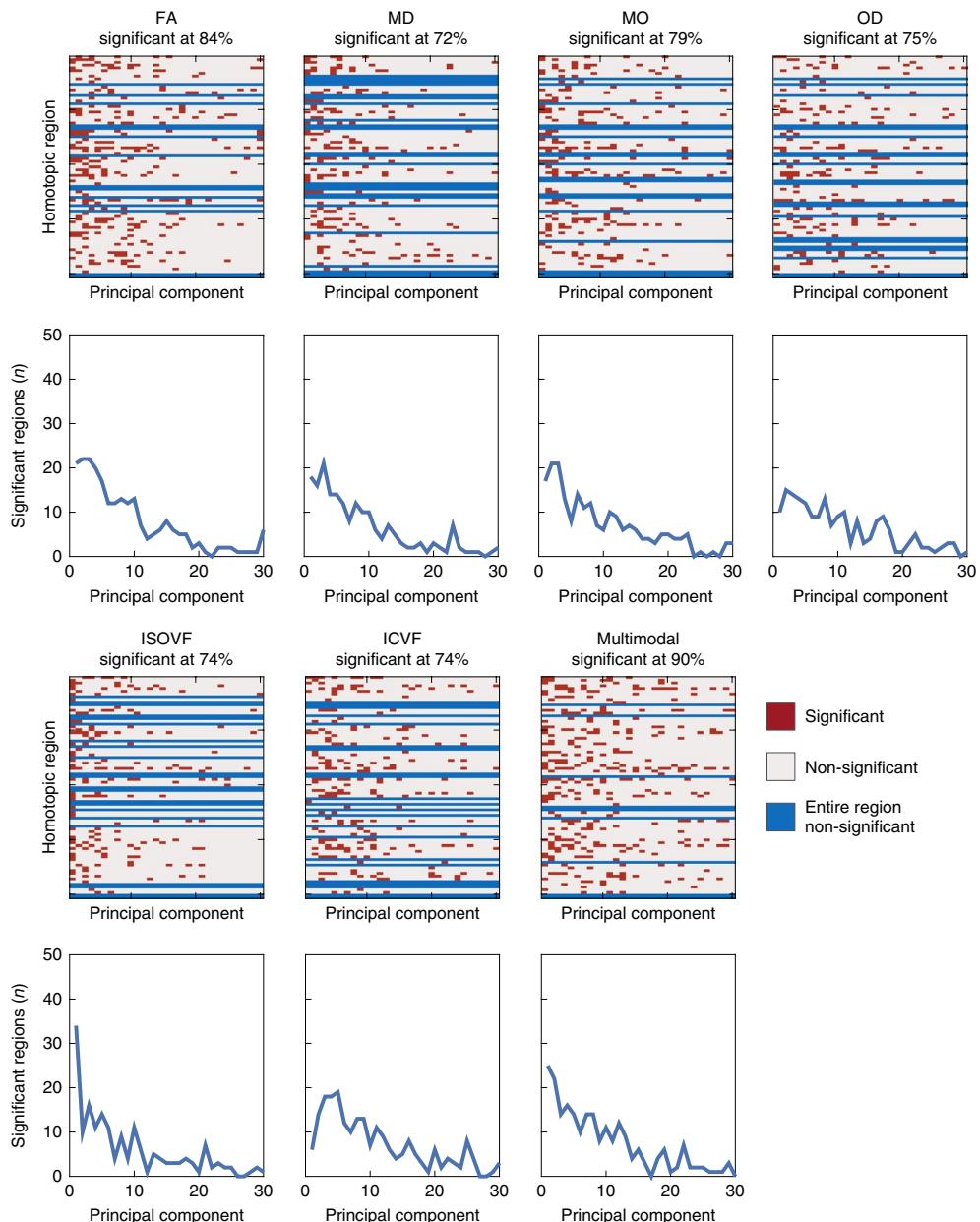


Fig. 3 | Significant associations between functional connectivity and microstructure of the connecting white matter tract. Each row in the matrices represents a homotopic region pair, with each entry a regressor (on the microstructural principal components) of the linear model. Significance of the regressors is color-coded. The graphs depict the number of regions for which a particular rank order principal component yielded a significant regressor (so, for example, in the multimodal models, the first principle component was significant for 28 brain regions). The percentage of homotopic region pairs demonstrating at least one significant regressor is given in the label of each matrix. Statistical significance was determined using permutation testing (two-sided, $n=100,000$ permutations, $P_{\text{uncorrected}} < 2.9 \times 10^{-6}$, $P_{\text{corrected}} < 0.05$, corrected for multiple comparisons). See text for definition of abbreviations.

significance ($P < 0.05$, corrected) is indicated per microstructural metric in Fig. 3, and in more detail in Supplementary Fig. 2 as Manhattan plots of corrected P values (family-wise error). The overall regression model was able to predict a statistically significant amount of cross-subject variance in 72–90% of the homotopic brain regions (depending on the dMRI metric). The multimodal microstructure model combining the six dMRI metrics provided a prediction of functional connectivity for the largest number of regions (72, representing 90% of the total brain areas considered). This result is not trivially guaranteed given that this model had the same number of regressors (30) as the other models. These results suggest a general relationship between microstructure and functional connectivity. We can further consider individual regressors (that is,

specific principal components). The statistically significant regressors generally correspond to the top principal components (left-most columns in Fig. 3). This indicates that the highest cross-subject modes of microstructural variation also explain the most cross-subject variation in functional connectivity. As the regressors reflect the primary modes of variation in the dMRI data but are used to model the fMRI data, this property is not trivially guaranteed. For some regions, no significant associations were found between homotopic functional connectivity and a given microstructure metric. The multimodal microstructure model again resulted in the largest number of significant regressors.

Having established that a microstructure–function link exists in most brain regions, we now consider the apparent strength of

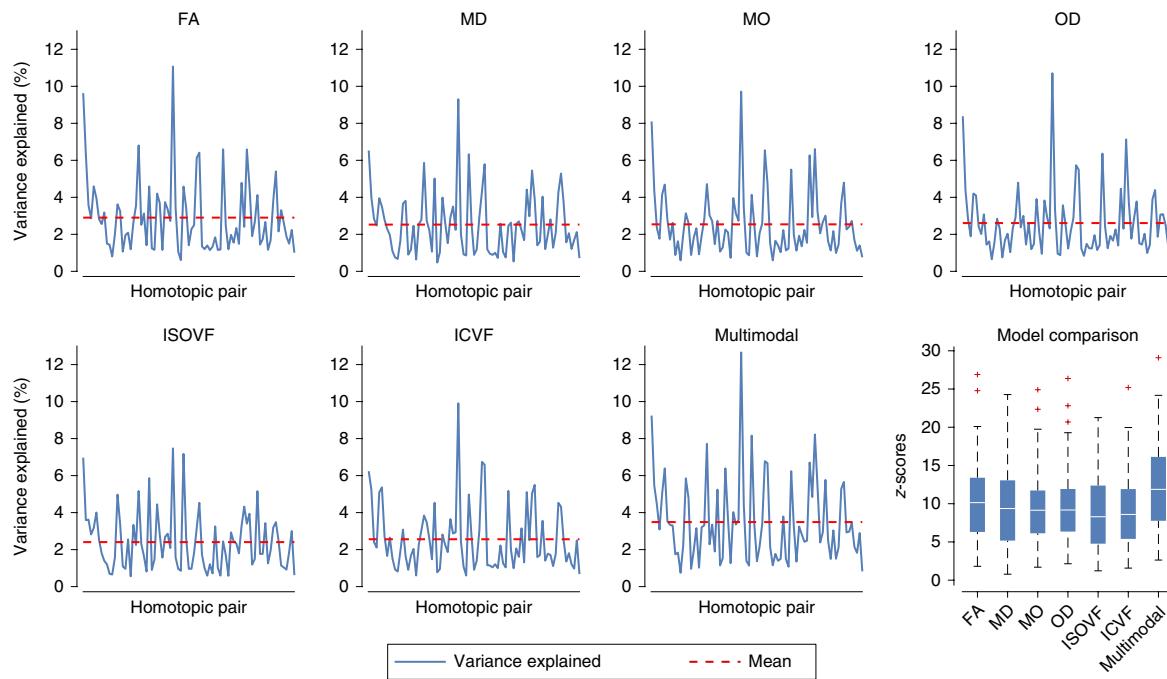


Fig. 4 | Percentage variance explained (r^2) in the functional connectivity of each homotopic region pair by microstructural metrics derived from the connecting white matter tract in the training cohort ($n=7,481$ subjects). The box-and-whisker plots at bottom right depict the model performance of each metric in terms of an f -to- z transformed score. The center line depicts the median z -scores across the homotopic region pairs; box limits are the 25th and 75th percentiles of the z -scores; whiskers extend to the most extreme data points excluding outliers (marked with a + symbol). See text for definition of abbreviations.

this relationship. Effect sizes of the regression models were evaluated in terms of percentage variance explained (equivalent to r^2) in functional connectivity by the microstructural metrics. The average variance explained across all significant model fits was 3.5% ($r=0.19$) for the multimodal model that combines all dMRI microstructure metrics. Substantial variation in variance explained was found across the different brain regions investigated (Fig. 4). In the multimodal regression, variance explained was lowest for the middle temporal gyrus (1.1%, $r=0.09$) and highest in the posterior cingulate cortex (12.7%, $r=0.36$). These effect sizes are mapped back to the 81 homotopic region pairs to visualize how strongly functional connectivity is explained by the underlying microstructure across the cortex (Fig. 5a). In addition, z -scores were computed to summarize the overall model fits. The multimodal microstructure regression model yielded on average a higher score than the regressions with any single microstructural metric ($z=12.0$, Fig. 4), suggesting that the different microstructural metrics explain different variance in functional connectivity. The model incorporating FA shows the highest average z -score of all individual metrics ($z=10.5$), although the different metrics perform overall fairly similarly (Fig. 4). A list of all brain areas investigated, with their corresponding effect sizes for the multimodal microstructure model, is given in Supplementary Table 1.

Tensor-based features (FA and MD in particular) have been shown to provide sensitive indicators of changes to tissue microstructure in a broad range of contexts. However, these measures can be influenced by multiple aspects of tissue microstructure¹², making interpretation difficult. We tested whether functional connectivity relates to a microstructure feature with greater biological specificity. We built on our previous work demonstrating quantitative agreement of OD estimates derived from dMRI data and with myelin stains in the same post-mortem human brain tissue²⁴. The callosal OD profile correlated well between the ex vivo imaging data (both MRI and microscopy) and the in vivo dMRI NODDI analyses

presented above, with both methods indicating high dispersion on the midline and lower dispersion in the lateral aspects of the corpus callosum (Supplementary Fig. 3c). Furthermore, OD estimates at the midline of the corpus callosum were able to explain significant variance in interhemispheric functional connectivity (Supplementary Fig. 3). While the explained variance (0.21% on average) was much less than with the spatially extended microstructure models presented above, the validation against histology demonstrates biological specificity of this particular association.

Model replication. We further tested the validity of the above models by applying them to the replication cohort of 3,873 subjects. The data for each replication subject were projected onto the 30 regressors and then multiplied by the regression coefficients estimated from the main cohort to predict that subject's functional connectivity. That is, the models were applied directly and not retrained on the new subjects. This therefore constitutes a direct prediction of functional connectivity from dMRI data in unseen subjects. As shown in Fig. 5, percentage variance explained was quantitatively very similar among regions (2.5% on average) in the previously unseen subjects, as in the main cohort upon which the model was based.

Several medial regions demonstrate notably high effect sizes, with the posterior cingulate cortex and the intra-calcarine cortex, in particular, having over 10% variance explained. Regions in the temporal lobe, ventral parts of the frontal lobe and lateral aspect of the occipital lobe demonstrate the lowest variance explained. In addition to the corpus callosum, temporal lobe regions are connected via the anterior commissure. For these regions, we performed additional analyses in which the microstructural signature from the anterior commissure was used to predict functional connectivity (see Supplementary Fig. 4). While the anterior commissure microstructure was able to predict functional connectivity, it did not explain the data better than callosal microstructure, nor did a model including both tracts.

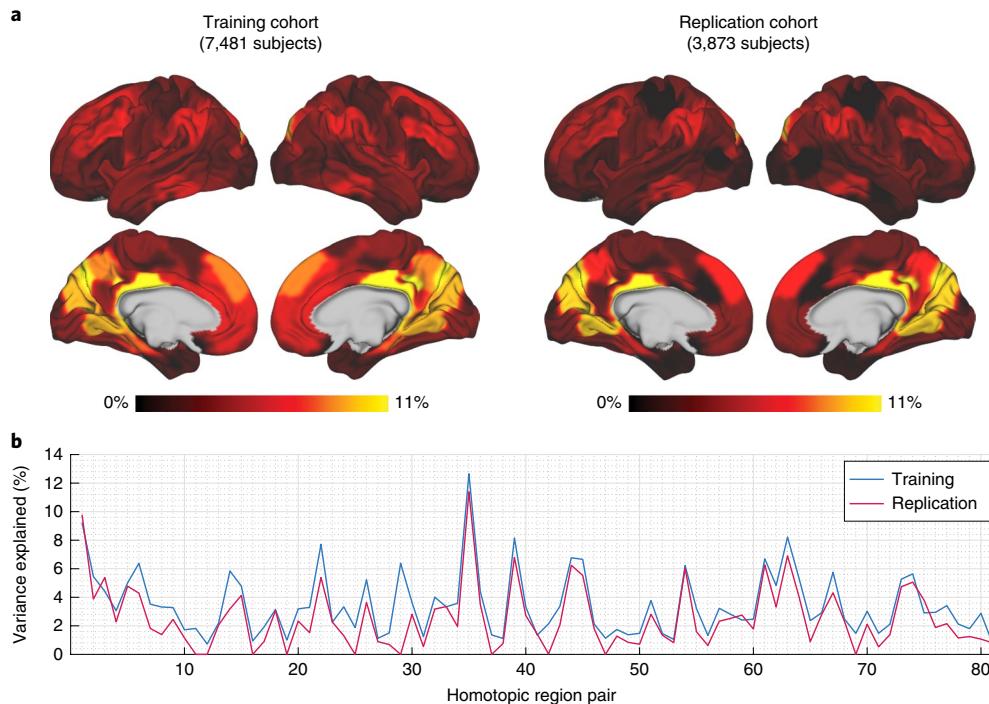


Fig. 5 | Total variance explained by the multimodal regression model in the training and replication cohorts. **a**, Variance explained mapped onto the brain surface. The maps were smoothed with a 2 mm Gaussian kernel to aid visualization. A similar pattern across the brain was found for the regression models incorporating the individual microstructural metrics. **b**, Graph showing percentage variance explained for each homotopic region. The model was trained on the main cohort of 7,481 subjects. By applying the regression models trained on the main cohort, we could predict functional connectivity in the replication cohort of 3,873 unseen subjects. The homotopic region numbers on the x axis correspond to the brain areas listed in Supplementary Table 1.

Negative control analysis. Although the above analyses suggest a general microstructure–function relationship, it is not clear whether these associations are specific to the pathway connecting a given pair of regions, or whether functional connectivity reflects global variance in the microstructural metrics across subjects. A new series of regression analyses were performed similar to those depicted in Fig. 2, but instead taking microstructure values from a different ‘wrong’ callosal tract (Fig. 6a). From the 81 callosal sub-regions defined above, we selected a subset of 30 distinct tracts with minimal spatial overlap (Supplementary Fig. 5) for use as control (‘wrong’) tracts (Fig. 6b). We then assessed whether any of the control tract regressions had similar or better performance compared to the correct tract (Fig. 6c). For 70% (60% in the replication cohort) of the homotopic areas, the highest z-score was obtained when the model was performed with the anatomically correct tract; overall, for 81% of brain areas the correct tract ranked among the top three models (Fig. 6d).

Genome-wide associations. We studied the influence of genetics on the microstructure–function relationships identified above with a series of GWASs. All subjects in this analysis were selected based on recent British ancestry and availability of genotype data that passed the quality control procedures of UK Biobank²⁵. The target phenotypes used in the GWASs were the cross-subject variation in functional connectivity predicted by the microstructure model (that is, the model fits; see Methods and Supplementary Fig. 6). For each homotopic region pair, the GWAS consisted of a series of univariate correlations of the model fit with 11,734,353 SNPs. These GWASs were fully multiple comparison corrected.

Fig. 7 depicts the association across SNPs for the homotopic pair with the largest variance explained in the multimodal microstructure model (that is, the posterior cingulate cortex). A group of SNPs in chromosome 14 demonstrated a strong association with the

microstructure–function phenotype. These SNPs were co-located with the gene Dishevelled Associated Activator of Morphogenesis 1 (*DAAM1*), while some were also within *DAAM1*’s promoter region (regulating expression of the gene)²⁶. The *DAAM1* protein plays an important role in the Wnt signaling pathway within the cell, indirectly regulating cell polarity and movement during development. In the central nervous system, this protein has been shown to facilitate the guidance of commissural axons at the embryonic stage in mice and *Drosophila*^{27,28}. Expression of the gene Jun N-terminal kinase 1-associated membrane protein (*JKAMP*) was also regulated by these SNPs, as demonstrated by 3D chromatin interaction data²⁹ (Virtual 4C³⁰). Furthermore, the GWAS revealed many SNPs within the gene lysophosphatidic acid receptor 1 (*LPARI*) located in chromosome 9. *LPARI* encodes one of the six receptors involved in the lysophosphatidic acid signaling pathway in the cell³¹. SNPs co-located with both *DAAM1* and *LPARI* were found for the microstructure–function association of multiple brain areas (Fig. 7 and Table 1). Detailed Manhattan plots at the location of *LPARI* and *DAAM1* are given in Supplementary Figs. 7 and 8, respectively. Manhattan plots depicting the GWAS for the microstructure–function model fits of each homotopic region pair in the discovery cohort can be found in Supplementary Fig. 9.

The GWAS was repeated for subjects in the replication cohort. Rather than using the model prediction approach described above, the multimodal microstructure models were first retrained to better explain functional connectivity with microstructure for these subjects (see Supplementary Fig. 10 for the effect of retraining). This approach was motivated to make the genetic replication analysis more fully independent of the discovery dataset. Replication GWAS was performed on microstructure–function phenotypes from the homotopic regions showing an association in chromosomes 9 and 14 in the original subjects. Following common practice for replication GWA studies³², only those SNPs that demonstrated a

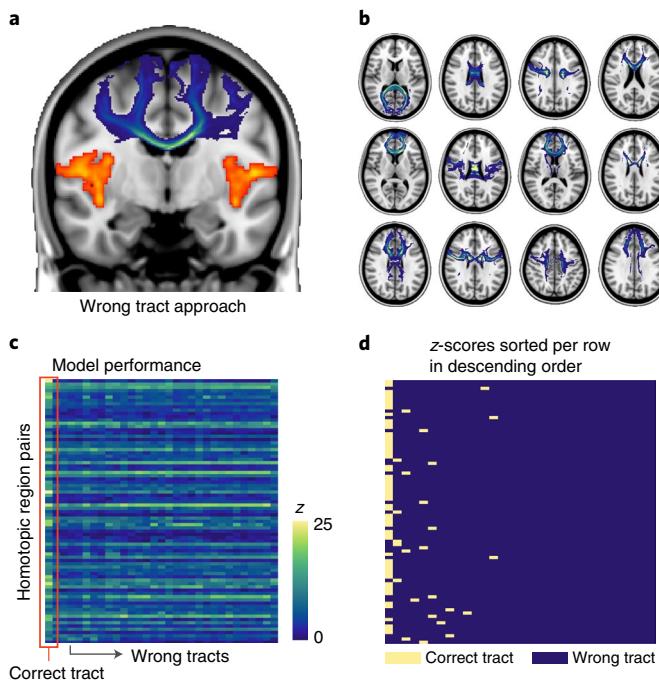


Fig. 6 | Negative control analysis. **a**, In the wrong tract approach, general linear model (GLM) analysis was performed with a tract (shown in blue) that does not directly connect between a homotopic pair of interest (shown in orange). **b**, A total of 30 distinct tracts (12 shown here) were chosen based on minimal spatial overlap between these. **c**, For each GLM, an f-statistic (across the whole model, with the degrees of freedom for model and error 30 and 7,450, respectively) was calculated and transformed to a z-score to compare the correct tract and 30 wrong tracts. All GLMs in this analysis were derived from the multimodal microstructural information. **d**, Rows from the matrix in **c** were sorted in descending order of z-score and labeled according to whether they represented the correct pathway or a different tract. The highest z-scores (left-most column) correspond to the anatomically correct tract in 70% of cases and, overall, the correct tract was in the top three models in 81% of cases.

significant association in the discovery GWAS were tested. For SNPs within the *LPA1* gene in chromosome 9, associations with three out of five brain areas were replicated. The SNPs in chromosome 14 corresponding to *DAAM1* were replicated in two out of three brain areas (Fig. 7).

The GWAS results described above used microstructure–function model fits as the target phenotype. These results could simply reflect correlations of these SNPs with both functional connectivity and microstructure. To test for specificity, two additional GWASs were run using the following target phenotypes for each homotopic pair: (1) the functional connectivity that remains unexplained by white matter microstructure (that is, the residuals from each multimodal microstructure–function model) and (2) the first principal component of the multimodal microstructure for the corresponding callosal pathway (that is, the first regressor in the microstructure model). SNPs that were significantly associated with either of these two phenotypes in the discovery and replication cohorts are listed in Supplementary Tables 2, 3. These GWASs did not find any SNPs co-located with either *DAAM1* or *LPA1* in any homotopic region. This suggests that the relationship to *DAAM1* and *LPA1* is specific to the component of functional connectivity that can be predicted by white matter microstructure. The GWAS associating with the first principal component of multimodal microstructure yielded SNPs within the *VCAN* gene, which were previously found to associate with ICVF throughout white matter in the brain¹⁶.

Discussion

Although basic principles relating axonal properties to neural signaling are well established, the degree to which functional connectivity is mediated by microstructural organization at the level of macroscopic tracts is largely unknown. Several studies have related the ‘strength’ and topology of structural connections to functional activity based on fMRI and dMRI^{10,33}, but these studies are uninformative about microstructure. Here we focused on commissural fibers passing through the corpus callosum, a set of connections that can be estimated using MRI both structurally and functionally. Our results are consistent with previous work^{17,18} in that connections between pairs of homotopic areas were shown to be the strongest functional connections in the brain. Furthermore, other studies have demonstrated that severing the corpus callosum reduces or extinguishes interhemispheric functional connectivity, providing evidence that communication between these regions is primarily facilitated by axons running through the callosum^{34,35}.

In this study, we have demonstrated that white matter microstructure is associated with functional connectivity at the macroscopic level probed by imaging. The majority of brain regions (90%) show statistical evidence for a relationship between white matter microstructure and functional connectivity. Replication in nearly 4,000 subjects demonstrates that the regression models fit in the main cohort have predictive power in unseen subjects.

On average these models account for 3–4% of the cross-subject variance in a given brain region, with considerable variation across regions—ranging from 1 to 13% of variance explained. It is likely that our results underestimate the true relationships due to methodological limitations. MRI provides indirect estimates of functional connectivity and microstructure. In addition, the model order (linear with 30 regressors), choice of confounds (Supplementary Table 4) and potential for remaining indirect connections (regions not included in the partial correlation) could all lead to unexplained variance. This may be one source of the inter-regional variation, although some true biological variation is also likely. Methodological improvements may well increase the strength of the observed effect sizes. The ability to identify subtle relationships in the order of 1% ($r=0.1$) is directly related to our large sample size; indeed, it is common for even smaller effect sizes to be considered valuable in genetic studies providing that replication is demonstrated. The recent advent of population-level imaging requires particular caution in distinguishing between the statistical significance and biological meaningfulness of a given result³⁶. Nevertheless, the identification of small effects can be a first step toward aggregate measures with greater explanatory power: for example, polygenic risk factors for disease combining univariate GWAS outcomes with small effect sizes have been enabled by population-level genetics studies³⁷.

In both the main and replication cohort, functional connectivity was best explained in regions close to the medial aspect of the brain—for example, the intra-calcarine and posterior cingulate cortices (see Fig. 5). It should be acknowledged that some of these regions emerged as single contiguous nodes after spatial ICA, where more distal homotopic pairs were separated by other brain structures (see Fig. 1). This could reflect fMRI signal blurring between the hemispheres, driving up the apparent functional connectivity. However, it is unclear why this functional connectivity would be better predicted by a completely independent measure of white matter microstructure estimated from diffusion MRI. To investigate this effect further, we conducted a correlation of white matter tract length with variance explained across homotopic region pairs, finding no significant relationship ($r=-0.04$, $P=0.70$).

For the majority (70%) of the homotopic pairs considered, the strongest model prediction was derived from microstructure in the anatomically correct pathway compared to microstructure obtained from any of the other 30 callosal pathways. This negative control analysis is informative because it establishes that

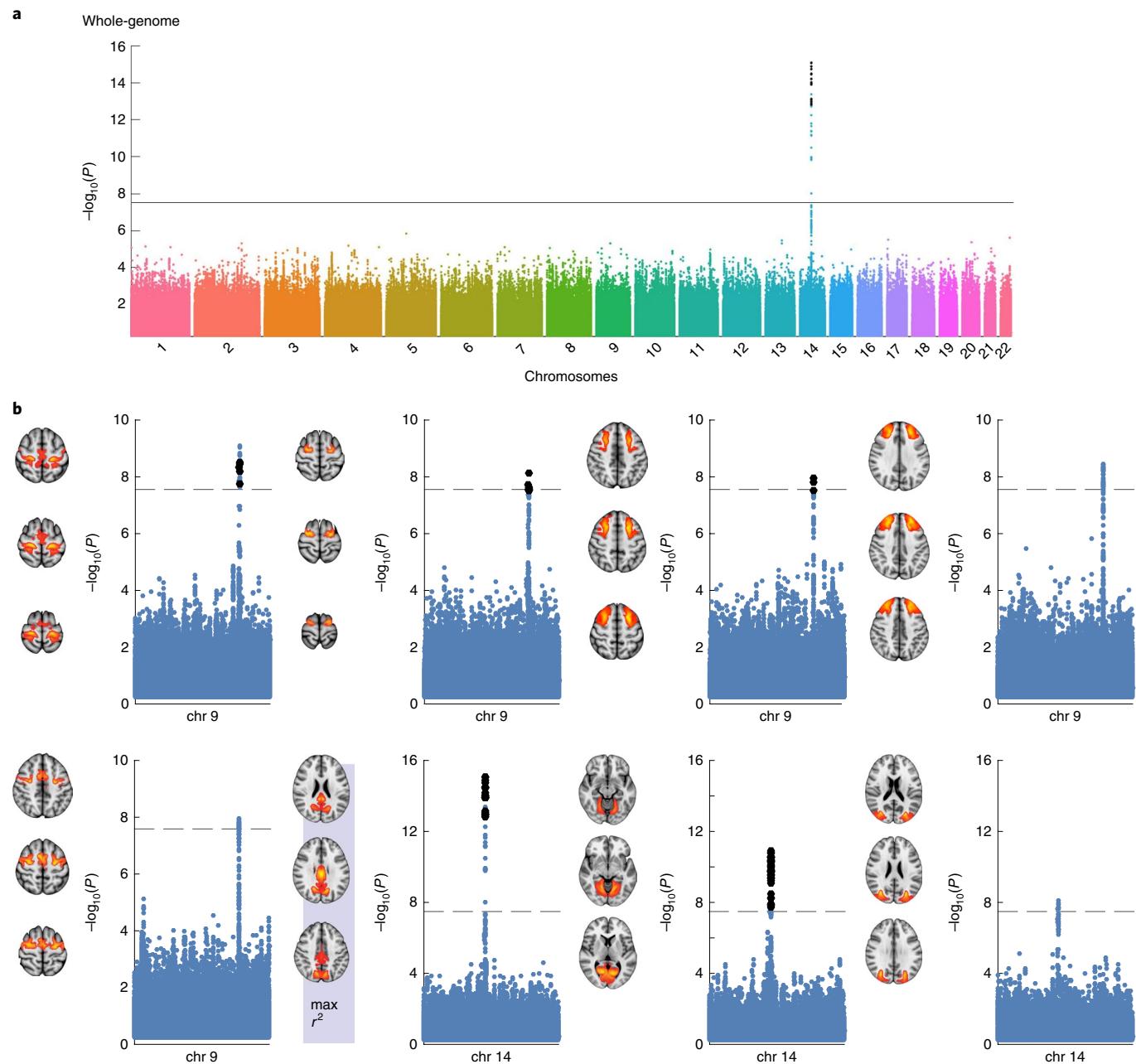


Fig. 7 | Genome-wide associations with the microstructure–function phenotype (that is, the pattern of functional connectivity that can be predicted from white matter microstructure). **a**, The Manhattan plot depicts the associations with each SNP across all chromosomes expressed as the $-\log_{10}(P)$ value. As an example, the genome-wide Manhattan plot is given for the homotopic brain region showing the highest variance explained by the microstructure–function model ($n = 7,481$ subjects). The strongest association is with an SNP (rs74826997) in chromosome 14 (linear regression, two-sided). **b**, Single-chromosome Manhattan plots are shown for brain regions that associate with SNPs in either chromosome 9 or 14 that co-located with the genes *LPA1* and *DAAM1*, respectively (linear regression, two-sided). The $-\log_{10}(P)$ values of the SNPs in the discovery GWAS (7,481 subjects) are depicted by blue dots. In an additional cohort of 3,873 subjects, we aimed to replicate the significant hits (black dots in single-chromosome Manhattan plots). The ICA spatial maps of these brain areas are given for each Manhattan plot. The brain area (posterior cingulate cortex, highlighted by max. r^2) corresponds to the genome-wide Manhattan plot in **a**. A significance threshold is given for a $-\log_{10}(P)$ value equal to 7.5 corresponding to a P value of $\sim 3 \times 10^{-8}$. The significance threshold for the replication GWAS was determined using Bonferroni correction ($P < 1.47 \times 10^{-4}$).

microstructure–function relationships have a degree of regional specificity and do not simply reflect global (brain-wide) inter-individual differences in microstructure and associated function. A similar result has been demonstrated for resting-state functional connectivity between the posterior cingulate and medial frontal cortices, with FA from the correct white matter pathway (cingulum) being more highly correlated than an unrelated tract¹³.

Interestingly, for a minority of the brain areas investigated, functional connectivity was better explained by microstructure from a ‘wrong’ white matter tract. Success of control tracts in predicting a given brain region could be driven by either confounded microstructural estimates in the correct tract, partial overlap of tract segmentations or global (brain-wide) variations in microstructure and functional connectivity.

Table 1 | Genome-wide associations (linear regression, two-sided) with the microstructure–function phenotype (that is, the pattern of functional connectivity that can be predicted from white matter microstructure).

Chr.	Node	RSID	Nearest gene	Function in central nervous system	Position	Ref. allele	Minor allele	Maf	Discovery P value	Replication P value
9	60	rs10980625	<i>LPAR1</i>	Lysophosphatidic acid signaling in central nervous system	113665018	C	C	0.11	7.24×10^{-9}	5.56×10^{-6}
	62	rs34860245	<i>LPAR1</i>		113709884	T	T	0.14	8.29×10^{-10}	3.76×10^{-5}
	67	rs4556147	<i>LPAR1</i>		113651161	A	A	0.22	1.09×10^{-8}	7.12×10^{-8}
14	1	rs76341705	<i>DAAM1</i>	Wnt signaling pathway, axonal growth and guidance	59628679	G	G	0.12	1.28×10^{-11}	7.15×10^{-8}
	35	rs74826997	<i>DAAM1</i>		59628609	T	T	0.12	8.15×10^{-16}	1.63×10^{-5}
15	33	rs1080066	<i>C15orf54</i>	Associated with spinal cord	39634222	A	A	0.09	3.48×10^{-13}	3.18×10^{-11}

Listed are the reference SNP cluster identifiers (RSID) of the SNPs showing the most significant associations replicated in the replication cohort. Some SNPs were associated with the microstructure–function model fits of multiple homotopic region pairs (highlighted in gray). The nearest gene of each SNP is reported with its possible function in the human central nervous system. Furthermore, the base-pair position, the SNP alleles, minor allele frequency (maf) and P values of the discovery ($n=7,481$ subjects) and replication GWAS ($n=3,873$ subjects) are given. A significance threshold is given for a $-\log_{10}(P\text{ value})$ equal to 7.5 corresponding to a P value of -3×10^{-8} . The significance threshold for the replication GWAS was determined using Bonferroni correction ($P < 1.47 \times 10^{-6}$).

A general and important confound in our models is ‘partial volume’ (spatial overlap) of tracts. Ideally one would estimate the properties of each axon connecting two brain regions. Instead, dMRI averages within a voxel or region of interest (in our case, a white matter tract). As such, dMRI measurements often mix multiple different white matter bundles: for example, in the centrum semiovale, callosal fibers cross the corticospinal tract and superior longitudinal fasciculus. As a result, our regression models will have included microstructural estimates from other tracts. More advanced modeling to exclude or model these partial volume effects would be valuable to increase specificity without reducing sensitivity.

Frontotemporal regions were particularly prone to being explained by a control tract. These regions also tend to have lower functional homotopic connectivity, in agreement with previous literature³⁸. Such regions may have fractionally less callosal input and be primarily connected to intra-hemispheric brain areas via associations fibers³⁹. Many temporo-polar regions also have interhemispheric connections via the anterior commissure. We therefore also constructed microstructural models from connections running through the anterior commissure; however, these models did not improve the explained variance in functional connectivity (see Supplementary Fig. 4). The effect sizes for models based on the anterior commissure and the corpus callosum varied similarly across brain regions. This may reflect either spatial overlap in the defined tracts close to gray matter, image registration errors or variation in non-relevant variance in functional connectivity across regions (providing a ceiling on the explainable variance). In addition to brain activity, other sources may also contribute to the resting-state signal (see refs. ^{40–42} for some excellent reviews).

Imaging microstructure with dMRI is a rapidly evolving field, including many models only recently developed. The biological interpretation of microstructural metrics is challenging and it is therefore not trivial to decompose the specific contributions of each microstructure parameter in explaining functional connectivity. Each microstructural metric used here explains some unique aspects of the dMRI signal, but they also share some mutual information (for example, both ICVF and OD correlate with FA). The results presented here demonstrate that combining these metrics yields a more comprehensive characterization of the underlying microstructure. In addition, having the rich representation of microstructure along the entire tract better explained functional connectivity than simply using the mean of the tract (Supplementary Fig. 11).

To gain further insight into the microscopic tissue features driving the dMRI-derived metrics, evaluation against reference

measures such as histology is essential. As such, we demonstrated good correspondence between OD profiles derived from the corpus callosum in ex vivo dMRI and myelin staining²⁴, providing confidence in the biological meaning of this specific measure. In agreement with histology^{24,43}, the dMRI data used in our study indicate that fibers are more dispersed at the center of the corpus callosum as compared to its lateral aspects. In Supplementary Fig. 3, we use this validated measure of fiber dispersion for a simple (single regressor) model to predict functional connectivity. However, this model provided much lower explanatory power (0.21% on average) than the multivariate regression models described above. This is probably because these more comprehensive models capture the spatial richness of microstructure metrics across the white matter tract, demonstrating how pooling of multiple white matter phenotypes can explain more variance in functional connectivity.

Data richness in the UK Biobank project allowed us to associate genetic variants with the imaging-derived phenotypes in this study. Meta-analyses in the ENIGMA consortium previously revealed genetic variants that were associated with imaging markers such as hippocampal volume⁴⁴ and other subcortical structures³² in over 30,000 subjects. ENIGMA pools a vast collection of imaging data from several studies acquired with heterogeneous protocols. It remains to be seen whether the inclusion of a large number of subjects in ENIGMA effectively mitigates this data heterogeneity. In contrast, the UK Biobank project aims to maximize data homogeneity in 100,000 subjects with a common protocol and imaging platform⁷, which may enable more efficient identification of associations between imaging phenotypes and genetic variants¹⁶. Combining datasets from the UK Biobank and ENIGMA can be used either to further boost statistical power or can be used separately to replicate discoveries.

We conducted a GWAS for each homotopic region pair to associate SNPs with the fraction of functional connectivity that was predicted by microstructure. In chromosomes 9 and 14, a group of SNPs was found showing a strong association with the cross-subject pattern of functional connectivity predicted by microstructure for multiple brain areas (Fig. 7). Because no SNPs associated with these same genes were found in GWASs relating solely to functional connectivity or microstructure, these associations appear to be unique to the microstructure–function relationship (see Supplementary Tables 2, 3). For the replication cohort, the SNPs in chromosome 9—co-located with *LPAR1*—were replicated for three of the five brain areas showing hits in the discovery GWAS. The SNPs in chromosome 14 were replicated in two out of the three brain areas.

The identified SNPs in chromosomes 9 and 14 are associated with genes that have previously been shown to be important for brain development. The *DAAM1* gene is expressed in many tissue of the human body and plays an important role in the Wnt signaling pathway⁴⁵. In neuronal tissue the *DAAM1* protein is primarily found in the shaft of neuronal dendrites⁴⁶, and in the developing brain it aids axonal guidance in targeting distal brain regions⁴⁷. Knockout studies in mice and *Drosophila* have shown deficits in the central nervous system when *DAAM1* is not expressed²⁷. In particular, the formation of commissural fibers at an embryonic stage was disturbed²⁸. Previous work relating cortical thickness to genetic variants also reported SNPs co-located with *DAAM1* in the cuneus area¹⁶ (<http://big.stats.ox.ac.uk>). Three-dimensional chromatin data revealed that the SNPs in chromosome 14 also regulate expression of the *JKAMP* gene²⁹. While diseases associated with *JKAMP* include medulloblastomas⁴⁸, its exact mechanism in brain development is not well described in the literature. For chromosome 9, several SNPs were located in the *LPAR1* gene, encoding a receptor involved in the lysophosphatidic acid signaling pathway. These receptors are found on the membranes of most cell types in the central nervous system and have been linked to certain neural processes including, but not limited to, neurogenesis, myelination, microglial activation and astrocyte responses^{31,49}.

The degree to which functional connectivity between brain regions is mediated by microscopic properties (microstructure) of the white matter pathways is a fundamental question in neuroscience. We demonstrated that a fraction of cross-subject variation in interhemispheric functional connectivity can be predicted from white matter tract microstructure connecting two homotopic regions. Our results suggest that microstructure–function relationships are general (across many brain regions), specific (with the correct tract out-predicting control tracts) and reproducible (as a prediction in a replication cohort). Furthermore, the microstructure–function association was underpinned by genetic variants and, in particular, with SNPs co-located with the genes *DAAM1* and *LPAR1*. Attribution of these relationships to specific biological sources, ideally in a causal manner, cannot be achieved with this kind of observational study but would probably require interventional studies in animals⁵⁰.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at <https://doi.org/10.1038/s41593-019-0379-2>.

Received: 5 November 2018; Accepted: 8 March 2019;

Published online: 15 April 2019

References

- Buzsáki, G. *Rhythms of the Brain* (Oxford University Press, 2006).
- Kötter, R. & Sommer, F. T. Global relationship between anatomical connectivity and activity propagation in the cerebral cortex. *Phil. Trans. R. Soc. Lond. B* **355**, 127–134 (2000).
- Rorden, C. & Karnath, H.-O. Using human brain lesions to infer function: a relic from a past era in the fMRI age? *Nat. Rev. Neurosci.* **5**, 813–819 (2004).
- Schmahmann, J. D. & Pandya, D. N. in *Fiber Pathways of the Brain* 1–37 (Oxford University Press, 2006).
- Basser, P. J., Pajevic, S., Pierpaoli, C., Duda, J. & Aldroubi, A. In vivo fiber tractography using DT-MRI data. *Magn. Reson. Med.* **44**, 625–632 (2000).
- Jbabdi, S. & Johansen-Berg, H. Tractography: where do we go from here? *Brain Connect.* **1**, 169–183 (2011).
- Miller, K. L. et al. Multimodal population brain imaging in the UK Biobank prospective epidemiological study. *Nat. Neurosci.* **19**, 1523–1536 (2016).
- Greicius, M. D., Supekar, K., Menon, V. & Dougherty, R. F. Resting-state functional connectivity reflects structural connectivity in the default mode network. *Cereb. Cortex* **19**, 72–78 (2009).
- van den Heuvel, M. P., Mandl, R. C. W., Kahn, R. S. & Hulshoff Pol, H. E. Functionally linked resting-state networks reflect the underlying structural connectivity architecture of the human brain. *Hum. Brain Mapp.* **30**, 3127–3141 (2009).
- Hagmann, P. et al. Mapping the structural core of human cerebral cortex. *PLoS Biol.* **6**, e159 (2008).
- Koch, M. A., Norris, D. G. & Hund-Georgiadis, M. An investigation of functional and anatomical connectivity using magnetic resonance imaging. *Neuroimage* **16**, 241–250 (2002).
- Beaulieu, C. The basis of anisotropic water diffusion in the nervous system - a technical review. *NMR Biomed.* **15**, 435–455 (2002).
- van den Heuvel, M., Mandl, R., Luigjes, J. & Hulshoff Pol, H. Microstructural organization of the cingulum tract and the level of default mode functional connectivity. *J. Neurosci.* **28**, 10844–10851 (2008).
- Wahl, M. et al. Human motor corpus callosum: topography, somatotopy, and link between microstructure and function. *J. Neurosci.* **27**, 12132–12138 (2007).
- Zhang, H., Schneider, T., Wheeler-Kingshott, C. A. & Alexander, D. C. NODDI: practical in vivo neurite orientation dispersion and density imaging of the human brain. *Neuroimage* **61**, 1000–1016 (2012).
- Elliott, L. T. et al. Genome-wide association studies of brain imaging phenotypes in UK Biobank. *Nature* **562**, 210–216 (2018).
- Stark, D. E. et al. Regional variation in interhemispheric coordination of intrinsic hemodynamic fluctuations. *J. Neurosci.* **28**, 13754–13764 (2008).
- Shen, K. et al. Stable long-range interhemispheric coordination is supported by direct anatomical projections. *Proc. Natl Acad. Sci. USA* **112**, 6473–6478 (2015).
- Alfaro-Almagro, F. et al. Image processing and quality control for the first 10,000 brain imaging datasets from UK Biobank. *Neuroimage* **166**, 400–424 (2018).
- Basser, P. J., Mattiello, J. & LeBihan, D. Estimation of the effective self-diffusion tensor from the NMR spin echo. *J. Magn. Reson.* **103**, 247–254 (1994).
- Ennis, D. B. & Kindlmann, G. Orthogonal tensor invariants and the analysis of diffusion tensor magnetic resonance images. *Magn. Reson. Med.* **55**, 136–146 (2006).
- Behrens, T. E. J., Berg, H. J., Jbabdi, S., Rushworth, M. F. S. & Woolrich, M. W. Probabilistic diffusion tractography with multiple fibre orientations: what can we gain? *Neuroimage* **34**, 144–155 (2007).
- Smith, S. M. et al. Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data. *Neuroimage* **31**, 1487–1505 (2006).
- Mollink, J. et al. Evaluating fibre orientation dispersion in white matter: comparison of diffusion MRI, histology and polarized light imaging. *Neuroimage* **157**, 561–574 (2017).
- Bycroft, C. et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
- Kent, W. J. et al. The human genome browser at UCSC. *Genome Res.* **12**, 996–1006 (2002).
- Matusek, T. et al. Formin proteins of the DAAM subfamily play a role during axon growth. *J. Neurosci.* **28**, 13310–13319 (2008).
- Avilés, E. C. & Stoeckli, E. T. Canonical wnt signaling is required for commissural axon guidance. *Dev. Neurobiol.* **76**, 190–208 (2016).
- Won, H. et al. Chromosome conformation elucidates regulatory relationships in developing human brain. *Nature* **538**, 523–527 (2016).
- Wang, Y. et al. The 3D Genome Browser: a web-based browser for visualizing 3D genome organization and long-range chromatin interactions. *Genome Biol.* **19**, 151 (2018).
- Yung, Y. C., Stoddard, N. C., Mirendil, H. & Chun, J. Lysophosphatidic acid signaling in the nervous system. *Neuron* **85**, 669–682 (2015).
- Hibar, D. P. et al. Common genetic variants influence human subcortical brain structures. *Nature* **520**, 224–229 (2015).
- Honey, C. J. et al. Predicting human resting-state functional connectivity from structural connectivity. *Proc. Natl Acad. Sci. USA* **106**, 2035–2040 (2009).
- O'Reilly, J. X. et al. Causal effect of disconnection lesions on interhemispheric functional connectivity in rhesus monkeys. *Proc. Natl Acad. Sci. USA* **110**, 13982–13987 (2013).
- Roland, J. L. et al. On the role of the corpus callosum in interhemispheric functional connectivity in humans. *Proc. Natl Acad. Sci. USA* **114**, 13278–13283 (2017).
- Smith, S. M. & Nichols, T. E. Statistical challenges in “big data” human neuroimaging. *Neuron* **97**, 263–268 (2018).
- Torkamani, A., Wineinger, N. E. & Topol, E. J. The personal and clinical utility of polygenic risk scores. *Nat. Rev. Genet.* **19**, 581–590 (2018).
- Tobey, S. M. et al. A surface-based technique for mapping homotopic interhemispheric connectivity: development, characterization, and clinical application. *Hum. Brain Mapp.* **37**, 2849–2868 (2016).
- Schmahmann, J. D. et al. Association fibre pathways of the brain: parallel observations from diffusion spectrum imaging and autoradiography. *Brain* **130**, 630–653 (2007).

40. Smith, S. M. The future of fMRI connectivity. *Neuroimage* **62**, 1257–1266 (2012).
41. Cordes, D. et al. Frequencies contributing to functional connectivity in the cerebral cortex in “resting-state” data. *Am. J. Neuroradiol.* **22**, 1326–1333 (2001).
42. Friston, K. J. Functional and effective connectivity: a review. *Brain Connect.* **1**, 13–36 (2011).
43. Budde, M. D. & Annese, J. Quantification of anisotropy and fiber orientation in human brain histological sections. *Front. Integr. Neurosci.* **7**, 3 (2013).
44. Hibar, D. P. et al. Novel genetic loci associated with hippocampal volume. *Nat. Commun.* **8**, 13624 (2017).
45. Habas, R., Kato, Y. & He, X. Wnt/Frizzled activation of Rho regulates vertebrate gastrulation and requires a novel Formin homology protein Daam1. *Cell* **107**, 843–854 (2001).
46. Salomon, S. N., Haber, M., Murai, K. K. & Dunn, R. J. Localization of the Diaphanous-related formin Daam1 to neuronal dendrites. *Neurosci. Lett.* **447**, 62–67 (2008).
47. Kida, Y., Shiraishi, T. & Ogura, T. Identification of chick and mouse Daam1 and Daam2 genes and their expression patterns in the central nervous system. *Brain Res. Dev. Brain Res.* **153**, 143–150 (2004).
48. Behrends, U. et al. Novel tumor antigens identified by autologous antibody screening of childhood medulloblastoma cDNA libraries. *Int. J. Cancer* **106**, 244–251 (2003).
49. González de San Román, E. et al. Anatomical location of LPA1 activation and LPA phospholipid precursors in rodent and human brain. *J. Neurochem.* **134**, 471–485 (2015).
50. Fields, R. D. A new mechanism of nervous system plasticity: activity-dependent myelination. *Nat. Rev. Neurosci.* **16**, 756–767 (2015).

Acknowledgements

All data in this study were obtained from the UK Biobank project (access no. 8107). We are very grateful to all individuals who donated their time to participate in the UK Biobank study. K.L.M., M.K. and J.Mo. are supported by the Wellcome Trust

(nos. 091509/Z/10/Z, 202788/Z/16/Z, 098369/Z/12/Z). The authors gratefully acknowledge funding from the Wellcome Trust UK Strategic Award (no. 098369/Z/12/Z). UK Biobank brain imaging and F.A.-A. are funded by the UK Medical Research Council and the Wellcome Trust. J.Ma. acknowledges funding for this work from the European Research Council (grant no. 617306) and the Leverhulme Trust. S.J. is supported by the UK Medical Research Council (no. MR/L009013/1). The Wellcome Centre for Integrative Neuroimaging is supported by core funding from the Wellcome Trust (no. 203139/Z/16/Z). Finally, we would like to thank D. Norris and J. Marques for their valuable input on this work.

Author contributions

J.Mo., S.M.S., S.J. and K.L.M. designed the research. J.Mo. performed the research. K.L.M., F.A.-A. and S.M.S., developed acquisition and processing pipelines for the MRI data. L.T.E. and J.Ma. processed genetics data, provided tools for genome-wide associations analysis and gave feedback on genetics results. J.Mo., S.M.S., M.K., M.H., A.M.C.W., S.J. and K.L.M. analyzed the data and interpreted its outcomes. J.Mo. and K.L.M. wrote the manuscript, which was edited by all authors.

Competing interests

J.Ma. is a co-founder and director of GENSCI Ltd. S.S. is a co-founder of SBGneuro.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41593-019-0379-2>.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to J.M.

Journal peer review information: Nature Neuroscience thanks Genevieve Konopka and other anonymous reviewer(s) for their contribution to the peer review of this work.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2019

Methods

Data acquisition and preprocessing. We used resting-state functional MRI and diffusion MRI data provided by the UK Biobank project. An extensive overview of the data acquisition protocols and image processing carried out on behalf of UK Biobank can be found elsewhere^{7,19}. Description of post-processing pipelines and acquisition protocols of MRI data in UK Biobank are available at http://biobank.ctsu.ox.ac.uk/crystal/docs/brain_mri.pdf. Unless stated otherwise, processing of the MR images was performed using FSL v5.0⁵¹. All imaging data were acquired on a 3 T Siemens Skyra MRI scanner (software platform VD13) using a 32-channel receiving head coil.

Resting-state fMRI data with 2.4 mm isotropic resolution and whole-brain coverage (field of view, $88 \times 88 \times 64$ matrix) were acquired in a 6 min session (multiband acceleration 8, repetition time (TR) = 0.735 ms, 490 time-points). The functional data were motion corrected⁵² and FIX-cleaned⁵³ to remove physiological noise and image artifacts, before transformation to a 2 mm MNI-template.

Diffusion MRI data were acquired at 2 mm isotropic resolution achieving whole-brain coverage (field of view, $104 \times 104 \times 72$ matrix) with two diffusion weightings (*b*-values) (*b* = 1,000, 2,000 s per mm^2), with 100 unique gradient directions over the two shells (50 directions per shell). The total acquisition time was 7 min (multiband acceleration 3, echo time (TE)/TR 92/3,600 ms). After eddy current correction of all images⁵⁴, tensor metrics (FA, MD, MO) were calculated from the lower shell (*b* = 1,000 s per mm^2) using DTIFIT. Both shells were used to estimate the NODDI model⁵⁵ metrics (ICVF, ISOVF, OD) using the AMICO toolbox⁵⁶.

While not explicitly used in this study, the UK Biobank imaging protocol includes several structural acquisitions that informed the quality control pipeline and served as registration references for the functional and diffusion data^{7,19}. T1-weighted structural scans were acquired using a 3D MPRAGE protocol ($1.0 \times 1.0 \times 1.0$ mm resolution, matrix $208 \times 256 \times 256$, inversion time (TI)/TR = 880/2,000 ms, in-plane acceleration 2). T2-weighted imaging using fluid-attenuated inversion recovery (FLAIR) contrast provided estimates of white matter hyperintensity (3D SPACE, $1.05 \times 1.0 \times 1.0$ mm resolution, $192 \times 256 \times 56$ matrix, TI/TR = 1,800/5,000 ms, in-plane acceleration 2).

Quality control. Quality control was applied at several stages in this study. First, all raw data were subject to a standard preprocessing pipeline⁵⁰ that generates several quality control measures. The starting point was the T1-weighted structural scan, which is essential for further processing of the other modalities (for example, the generation of brain masks, tissue segmentations and as a reference for registration). Subjects were excluded if registration to standard space failed, probably due to excessive head motion, atypical structure and/or anatomical abnormalities (for example, large ventricles). The full list of quality control measures derived from the T1-weighted images is given elsewhere¹⁹. Based on the T1-weighted anatomical image, 98% of all subjects were deemed suitable for further analysis. Next, the volume of white matter hyperintensities, used as confound variable, derived from the T2-FLAIR images was characterized with BIANCA⁵⁶. This feature detects atypical structures and individuals with overt pathology¹⁹. Subjects could additionally be excluded from further analysis on the basis of their dMRI and fMRI data due to bad echoplanar imaging (EPI) distortions, failed registration to T1, extreme bias fields, unusable fieldmaps and/or severe motion artifacts; 87% of the dMRI datasets and 94% of the rfMRI datasets were considered suitable for further analysis based on these quality control measures. Exclusion of the relatively large number of dMRI scans was caused by a change in processing protocol that deemed some early scans unsuitable.

All subjects selected in this study had both usable dMRI and rfMRI data in addition to suitable genetics data (see section UK Biobank genetics data, below, for more information). This yielded a total of 11,354 subjects: 7,481 in the main cohort and 3,873 in the replication cohort (randomly assigned). Overall, 5,393 females were included, the mean age was 62.8 (s.d. 7.4) years and all subjects had recent British ancestry. No power calculation was needed in advance and we used all available samples. UK Biobank is an observational prospective epidemiological study, and all analyses in our study used all available subjects that fulfilled the criteria described above. Hence there is no equivalent process of randomization that comes into this analysis. For the exact same reason, no blinding step was involved.

Variations in white matter microstructure and/or functional connectivity may be influenced by certain quality control measures (for example, head motion) in a subtle way that does not require subject exclusion, but which could confound associations. A set of variables of no interest (confounds) are listed in Supplementary Table 4 that were used to deconfound the data before modeling (see section Predicting functional connectivity from white matter microstructure, below).

fMRI processing. The resting-state fMRI data were fed into an ICA using the MELODIC tool⁵⁷ to identify resting-state networks present on average in the whole population. First, data were reduced to 100 dimensions using PCA and then fed into spatial ICA, from which 55 components corresponded to functional regions while the remaining 45 were judged to reflect physiological noise or image artifacts ('noise')^{7,19}. A functional component was split if it consisted of

non-contiguous brain regions, yielding 81 bilateral (homotopic) regions that were further split between the hemispheres to estimate interhemispheric connectivity (see Supplementary Table 1). Average time-series were generated for all ICA components (that is, homotopic areas and noise components) by a spatial regression of the subject's voxel-wise resting-state fMRI time-series with the ICA spatial maps. Further analyses were performed using the FSLNets toolbox (<https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FSLNets>). The average time-series within a homotopic area was demeaned and 'cleaned' by regressing out the time-series from the 45 'noise' component time courses. Functional connectivity was estimated between all pairs of components (2×81) by means of partial correlation of the cleaned time-series using Ridge regression with a regularization factor $\rho = 1$. Partial correlation aims to estimate direct connectivity between two areas by first regressing out all other regions' time-series before calculating the correlation (that is, established through inversion of the covariance matrix).

dMRI tractography. White matter tracts between functional regions were delineated using tractography. Up to three fiber orientations were fitted at each dMRI voxel in a Bayesian approach using bedpostX⁵⁸ modified for multi-shell data⁵⁹. Probabilistic tractography was then performed with the probtrackx2 algorithm²² by generating streamlines from a seed region (5,000 voxel⁻¹) in one hemisphere and saving only streamlines that passed through the corpus callosum and terminated in the same region in the contralateral hemisphere. This process was repeated by switching the seed and the target area between hemispheres. The overlap of the identified tracts in this two-step approach was used to generate the mask corresponding to the tract of interest. The tracts were generated for all 81 homotopic pairs (each representing either the seed or the target area) for ten subjects drawn from the UK Biobank dataset. Tracts between a given homotopic pair were then averaged across these subjects and served as a tract mask for all subjects stored in 1 mm MNI-space.

Tract-based spatial statistics. Tract-based spatial statistics²³ was used to align white matter tracts between subjects and extract microstructural information from the tract center (skeleton). The version of TBSS used here employs an optimized nonlinear registration (FNIRT) that avoids the need for the projection step in the original version of TBSS⁶⁰. This avoids misalignment problems in which voxels can be projected onto a different tract that is in close proximity, an issue that has been highlighted from the original method⁶¹. The choice of FNIRT-based registration was motivated by its performance compared to other registration algorithms, as described previously¹⁹. We also evaluated the use of DTITK registration, which incorporates the full diffusion tensor to further improve the alignment of dMRI scans⁶², finding equivalent performance between the two algorithms (Supplementary Figs. 12, 13). The tract reconstructions obtained with probabilistic tracking were used to mask the white matter skeleton voxels for a given homotopic region pair. Microstructural features derived from the diffusion tensor and NODDI fits were extracted from this final tract mask.

Predicting functional connectivity from white matter microstructure. We used a multiple linear regression model to predict homotopic functional connectivity from a set of regressors describing the spatial pattern of microstructure along a white matter tract. A rank-based inverse normal transformation was applied to all data to ensure normality. The regression model was constructed separately for each pair of homotopic regions:

$$Y_i = X_i\beta + \epsilon_p \quad \text{with } i = 1, \dots, n$$

Here Y_i ($N_{\text{subjects}} \times 1$) is a vector that contains the functional connectivity values of all subjects derived from homotopic region i (over $n = 81$ regions). To build a model using p microstructural regressors, we need to estimate a set of regression coefficients β ($p \times 1$) that describe the relative contribution from the microstructural metrics X_i ($N_{\text{subjects}} \times p$) along the white matter tract.

The regressors were derived in two stages. First, the microstructural metrics were extracted from the TBSS voxels (white matter skeleton) corresponding to the tract of interest for every subject, yielding a matrix X^o_i ($N_{\text{subjects}} \times N_{\text{voxels}}$). As the matrix X^o_i is very large, a direct regression with functional connectivity is ill-conditioned. We therefore performed a dimensionality reduction on X^o_i to derive a set of regressors reflecting the primary modes of variation of a given microstructural metric across space for the cohort of subjects. The microstructural matrices were first demeaned, and then a singular value decomposition was computed from matrix X^o_i . The top p components were retained, yielding matrix X_i ($N_{\text{subjects}} \times p$). In practice, p was set to 30 principal components that approximately corresponded to a transition in the spectrum of singular values in terms of variance explained. This provides a somewhat conservative model order below the point around $p = 100$, at which variance explained roughly tracked noise singular vectors (Supplementary Fig. 1) and linear regression is prone to overfitting.

Matrices X_i were constructed for each of the microstructure metrics separately, yielding six single-metric linear regression models per homotopic region. In addition, a multimodal regression model was created that combined across all microstructure metrics. For the multimodal regression, all raw microstructure matrices (X^o_i) were demeaned and normalized through division by their first

singular value to ensure a comparable range of values. The six normalized matrices were then concatenated along the voxel dimension ($N_{\text{subjects}} \times 6N_{\text{voxel}}$), and this matrix was reduced to the top 30 principal components as described above.

We defined a set of 64 confound variables of no interest that might bias the estimated regressors by correlating with the estimated microstructural measures (for example, through artifacts such as partial volume). An overview of all confound variables is given in Supplementary Table 4. The confound variables were regressed out of the functional and microstructural data before fitting the regression models.

Statistical analysis. Statistical significance of the regression models was assessed by means of permutation testing, evaluating each regressor using a *t*-statistic. A null distribution was constructed for each model *t*-statistic by randomly permuting the functional connectivity values across subjects (100,000 permutations). Because multiple models were evaluated, correction for family-wise error is also essential, where we corrected along three different dimensions of multiple comparison, as follows. First, we tested multiple hypotheses in each model, that is, which of the 30 microstructural principal components explained a significant amount of functional connectivity. Second, the models were applied to each of the 81 homotopic region pairs. Finally, a total of seven models (six individual microstructural models and one multimodal model) were evaluated for each homotopic pair. Following the approach demonstrated in ref. ⁶², a maximum *t*-statistic null distribution across all dimensions (regressor, regions and models) was generated from the permuted *t*-statistics. From this maximum *t*-statistics null distribution, a corrected *P* value was estimated for each of the non-permuted *t*-statistics. Furthermore, an *f*-statistic was computed to judge the overall performance of each regression model (degrees of freedom model and error, 30 and 7,450, respectively). The *f*-statistics were converted to *z*-scores. Finally, the effect size of the regression models was expressed in terms of percentage variance explained (equivalent to *r*²), describing the strength of the relationship between microstructure and functional connectivity.

Negative control analysis. The statistical tests described above determine whether there is a relationship between functional connectivity in a given brain region and the microstructure in the white matter pathway that connects them. However, this does not provide any insight into whether these relationships are specific: for example, microstructure and function could correlate at the whole-brain level. In this case, a regression model could indicate a statistically significant relationship even when using a white matter pathway that does not connect a given homotopic pair. Such a relationship could still be biologically meaningful, but the interpretation would change (for example, demonstrating that individual brains vary globally from hypo- to hyperconnected).

To test this, a negative control analysis was performed to evaluate the uniqueness of the microstructure–function relationships. From the 81 tracts in our study, a subset of 30 with minimal mutual overlap were selected as canonical control ('wrong') tracts. To identify the set of canonical control tracts, the Dice similarity index was computed among all tracts to quantify spatial overlap. Using *k*-means clustering (*k*=3 clusters), a cluster of tracts with the lowest average similarity indices was selected (Supplementary Fig. 5).

The regression models were then re-evaluated for each homotopic area using the control tracts, rather than microstructure from the anatomically correct tract, for the homotopic pair of interest. If, for a homotopic area, the anatomically correct tract was among the control tracts, an additional control tract was selected. To summarize, the regression models of the homotopic regions were performed once for microstructure from the correct tract and 30 times for the control tracts. Comparison between the correct and control tract analyses was conducted using the *f*-statistic converted to *z*-scores.

UK Biobank genetics data. The GWASs were performed using the BGENIE software²⁵. Acquisition and processing steps of the genetics dataset for all subjects in the UK Biobank project can be found in ref. ²⁵. For the discovery cohort, we began with the set of 12,623 brain-imaged UK Biobank subjects for whom data were released in July 2017. As in ref. ¹⁶, to avoid confounding effects that might arise from population structure or environmental effects, we selected a subset of 11,354 unrelated subjects with recent British ancestry. Ancestry was determined using sample quality control information provided by UK Biobank²⁵. We then filtered the genetic data to remove SNPs with minor allele frequency <0.01% or a Hardy–Weinberg equilibrium *P* value of <10⁻⁷, yielding a total of 11,734,353 SNPs distributed across the 22 autosomes. Not all of the UK Biobank subjects who underwent brain imaging have usable data with a given MRI modality. All 11,354 unrelated samples were subjects which had usable dMRI and fMRI data according to previous quality control¹⁹. Subjects were assigned to the discovery and replication cohorts in a fashion similar to the MRI analyses.

Ex vivo MRI and histology data. Microscopy and MRI data from three ex vivo corpus callosum specimens were acquired and processed as described previously²⁴. In brief, formalin-fixed human brain tissue sections were scanned on a preclinical 9.4 T Varian MRI system. Diffusion MRI was performed by spin-echo sequence with TE=29 ms and TR=2.4 s. Two shells were acquired ($b=2,500$ s per mm²

and $b=5,000$ s per mm²), each with 120 gradient directions and 0.4 mm isotropic resolution. Eight images with no diffusion weighting were acquired. A parametric model was fit to the dMRI signals from the $b=5,000$ s per mm² dataset to obtain orientation dispersion estimates⁶³.

Following MR scanning, the specimens were histologically sectioned and immunohistochemically stained for myelin (proteolipid protein). The sections were digitized and we obtained fiber orientation estimates at each pixel using structure tensor analysis⁶⁴. From a two-dimensional local neighborhood (0.4×0.4 mm) corresponding to the size of an MRI voxel, a fiber orientation distribution was computed from which orientation dispersion was derived. After registration of dMRI and microscopy data to the same image space⁶⁵, dispersion estimates were compared against each other in the corpus callosum.

Ethics and informed consent. All participants in the UK Biobank project signed an informed consent, which is controlled by a dedicated Ethics and Guidance Council (<http://www.ukbiobank.ac.uk/ethics>). The Ethics and Governance Framework can be found at <http://www.ukbiobank.ac.uk/wp-content/uploads/2011/05/EGF20082.pdf>. Institutional Review Board approval, also from the North West Multi-center Research Ethics Committee, was obtained for the Ethics and Governance Framework.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

All source data (including raw and processed brain imaging data and genetics data) are available from UK Biobank via their standard data access procedure (see <http://www.ukbiobank.ac.uk/register-apply>).

Code availability

The image processing pipelines of the MRI data in the UK Biobank project can be found at <http://www.fmrib.ox.ac.uk/ukbiobank>. Custom-written Matlab code including the microstructure–function modeling is freely available at <https://users.fmrib.ox.ac.uk/~jmollink/Biobank/Biobank.html>.

References

51. Jenkinson, M., Beckmann, C. F., Behrens, T. E. J., Woolrich, M. W. & Smith, S. M. FSL. *Neuroimage* **62**, 782–790 (2012).
52. Jenkinson, M., Bannister, P., Brady, M. & Smith, S. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage* **17**, 825–841 (2002).
53. Salimi-Khorshidi, G. et al. Automatic denoising of functional MRI data: combining independent component analysis and hierarchical fusion of classifiers. *Neuroimage* **90**, 449–468 (2014).
54. Andersson, J. L. R. & Sotiroopoulos, S. N. An integrated approach to correction for off-resonance effects and subject movement in diffusion MR imaging. *Neuroimage* **125**, 1063–1078 (2016).
55. Daducci, A. et al. Accelerated microstructure imaging via convex optimization (AMICO) from diffusion MRI data. *Neuroimage* **105**, 32–44 (2015).
56. Griffanti, L. et al. BIANCA (Brain Intensity AbNormality Classification Algorithm): a new tool for automated segmentation of white matter hyperintensities. *Neuroimage* **141**, 191–205 (2016).
57. Beckmann, C. F. & Smith, S. M. Probabilistic independent component analysis for functional magnetic resonance imaging. *IEEE Trans. Med. Imaging* **23**, 137–152 (2004).
58. Behrens, T. E. J. et al. Characterization and propagation of uncertainty in diffusion-weighted MR imaging. *Magn. Reson. Med.* **50**, 1077–1088 (2003).
59. Jbabdi, S., Sotiroopoulos, S. N., Savio, A. M., Graña, M. & Behrens, T. E. J. Model-based analysis of multishell diffusion MR data for tractography: how to get over fitting problems. *Magn. Reson. Med.* **68**, 1846–1855 (2012).
60. de Groot, M. et al. Improving alignment in tract-based spatial statistics: evaluation and optimization of image registration. *Neuroimage* **76**, 400–411 (2013).
61. Bach, M. et al. Methodological considerations on tract-based spatial statistics (TBSS). *Neuroimage* **100**, 358–369 (2014).
62. Winkler, A. M., Ridgway, G. R., Webster, M. A., Smith, S. M. & Nichols, T. E. Permutation inference for the general linear model. *Neuroimage* **92**, 381–397 (2014).
63. Sotiroopoulos, S. N., Behrens, T. E. J. & Jbabdi, S. Ball and rackets: inferring fiber fanning from diffusion-weighted MRI. *Neuroimage* **60**, 1412–1425 (2012).
64. Budde, M. D. & Frank, J. A. Examining brain microstructure using structure tensor analysis of histological sections. *Neuroimage* **63**, 1–10 (2012).
65. Heinrich, M. P. et al. MIND: modality independent neighbourhood descriptor for multi-modal deformable registration. *Med. Image Anal.* **16**, 1423–1435 (2012).

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

Data was not explicitly acquired in this study, but originated from the UK Biobank project. Details about genetics data acquisition can be found at Bycroft et al 2017. Genotyping of all subjects was performed using the Applied BiosystemsTM UK BiLEVE AxiomTM Array by Affymetrix1 (807,411 markers) or using the closely-related Applied BiosystemsTM UK Biobank AxiomTM Array (825,927 markers). Both arrays were purpose-designed specifically for the UK Biobank genotyping project and share 95% of marker content.

MRI data acquisition and protocols can be found in Alfaro-Almagro et al., 2018 and Miller et al., 2016 or at http://biobank.ctsu.ox.ac.uk/crystal/docs/brain_mri.pdf. All MRI data were acquired with 3T Siemens Skyra (software platform VD13).

Data analysis

FMRIB's Software Library (FSL) v5.0 for MRI (pre)processing.
Part of FSL:

- BIANCA for estimating white matter hyperintensities from T2-FLAIR data
- MELODIC tool for estimating resting-state networks from fMRI data using ICA
- FSLNets toolbox for processing the resting-state time-series and network analyses
- BedpostX for estimating sub-voxel fibre configurations from diffusion MRI data.
- Protrackx2 to delineate white matter bundles using tractography.
- TBSS for estimating the tract skeleton of white matter bundles.

AMICO for estimating NODDI parameters from diffusion MRI.
Matlab r2016a for data analysis, multivariate modeling and plotting.
BGENIE v1.2 for genome-wide associations.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The UK Biobank Brain imaging protocol consists of 6 distinct modalities covering structural, diffusion and functional imaging, summarised in Supplementary Table 1. For this study, we primarily used data from the February 2017 release of ~10,000 participants' imaging data (and an additional ~5,000 subjects' data released in January 2018 provided the larger replication sample). The raw data from these 6 modalities has been processed for UK Biobank to create a set of imaging derived phenotypes (IDPs). These are available from UK Biobank, and it is these IDPs from the 2017/18 data releases that we used in this study (see <http://www.ukbiobank.ac.uk/register-apply>).

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No sample-size calculation was performed, as large epidemiological studies like the UK Biobank project (and others) aim to maximize sample sizes to detect small variations across subjects. The sample sizes were determined based on the availability of usable MRI data in addition to genetics data without artefacts from subjects with a recent British ancestry. These sample sizes deemed sufficient to detect effect sizes reported in our work (>0.1% variance explained).
Data exclusions	<p>Exclusion criteria were pre-established. With regards to the MRI data, a subject can be excluded based on the T1-weighted scan if registration to standard space fails, likely due to excessive head motion, atypical structure and/or anatomical abnormalities (e.g., large ventricles). Subjects can additionally be excluded from further analysis on the basis of their dMRI and fMRI data due to bad EPI distortions, failed registration to T1, extreme bias fields, unusable fieldmaps and/or severe motion artefacts. 87% of the dMRI datasets and 94% of the rfMRI datasets were considered suitable for further analysis based on these QC measures.</p> <p>Similarly, subjects were selected on usable genetics data. As in Elliott et al, to avoid confounding effects that may arise from population structure or environmental effects, we selected unrelated subjects with recent British ancestry. Ancestry was determined using sample quality control information provided by UK Biobank. We then filtered the genetic data to remove SNPs with minor allele frequency < 0.01% or a Hardy-Weinberg equilibrium p-value of less than 10⁻⁷, yielding a total of 11,734,353 SNPs distributed across the 22 autosomes. All subjects used in our study passed these quality control criteria, yielding a total of 11354 subjects</p>
Replication	All models estimating functional connectivity (derived from resting-state fMRI) from white matter microstructural metrics (derived from diffusion MRI) were trained on the main cohort of 7481 subjects. We used an additional subset of 3873 subjects to replicate these findings. Likewise, we performed a genome-wide association in the main cohort and replicated the reported genetic variants in the replication cohort.
Randomization	Subjects were randomly assigned to the two cohorts (main and replication).
Blinding	There were no experimental groups, so no blinding steps were involved in this study.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
<input checked="" type="checkbox"/>	Antibodies
<input checked="" type="checkbox"/>	Eukaryotic cell lines
<input checked="" type="checkbox"/>	Palaeontology
<input checked="" type="checkbox"/>	Animals and other organisms
<input type="checkbox"/>	Human research participants
<input checked="" type="checkbox"/>	Clinical data

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	ChIP-seq
<input checked="" type="checkbox"/>	Flow cytometry
<input type="checkbox"/>	MRI-based neuroimaging

Human research participants

Policy information about [studies involving human research participants](#)

Population characteristics	Sample size, 11354 subjects (7481 subjects in the main cohort + 3873 subjects in the replication cohort); 5393 females; age, 62.8 (SD 7.4) years. All subjects had recent British ancestry.
Recruitment	Participants were selected using the NHS register, and invited to volunteer for the study. Recruitment was carried out between 2007 and 2010. Full details of the recruitment process are available in UK Biobank: Protocol for a large-scale prospective epidemiological resource, 2007 (http://www.ukbiobank.ac.uk/wp-content/uploads/2011/11/UK-Biobank-Protocol.pdf)
Ethics oversight	Ethics is handled by the UK Biobank Ethics Advisory Committee. Further details can be found at https://www.ukbiobank.ac.uk/ethics/

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Magnetic resonance imaging

Experimental design

Design type	Resting-state functional MRI data, no task.
Design specifications	No tasks were performed during fMRI scanning, so no design.
Behavioral performance measures	No tasks were performed during fMRI scanning, so no behavioral performance measures.

Acquisition

Imaging type(s)	T1-weighted MRI, T2-weighted MRI, diffusion MRI and resting-state functional MRI
Field strength	3T
Sequence & imaging parameters	<p>T1-weighted MRI: 3D MPAGE protocol TI = 880 ms, TR = 2000 ms and an in-plane acceleration factor of 2. Field of view 208x256x256 matrix Resolution 1.0x1.0x1.0 mm</p> <p>T2-weighted MRI: Fluid-attenuated inversion recovery (FLAIR) protocol (3D SPACE). TI = 1800, TR = 5000 ms and an in-plane acceleration factor of 2. Field of view: 192x256x56 matrix. Resolution: 1.05x1.0x1.0 mm.</p> <p>Diffusion MRI: Diffusion weighted spin-echo EPI sequence using multi-band (MB) acceleration. MB = 3, R = 1, TE/TR = 92/3600 ms, no iPAT, PF 6/8, fat saturation. Field of view: 104x104x72 matrix Resolution: 2x2x2mm. b = 0 (5x + 3x phase-encoding reversed), b = 1000, (50x), b = 2000 (50x).</p> <p>Resting-state functional MRI: Gradient echo EPI sequence using multi-band acceleration. TE/TR = 39/735 ms, MB = 8, R = 1, no iPAT, flip angle 52°, fat saturation. Field of view: 88x88x64 matrix. Resolution: 2.4x2.4x2.4mm. 490 time-points.</p>
Area of acquisition	Whole brain
Diffusion MRI	<input checked="" type="checkbox"/> Used <input type="checkbox"/> Not used
Parameters	100 gradient directions over two shells, 50 directions/shell. b-values = 1000, 2000 s/mm ²

Preprocessing

Preprocessing software	FMRIB's Software Library (FSL), v5.0
Normalization	Diffusion MRI: Subject's fractional anisotropy (FA) maps (obtained after fitting the diffusion tensor model to the raw diffusion MRI data) were transformed using FNIRT (part of FSL) to a 1-mm FA template in MNI-space.
Normalization template	Diffusion MRI: FA template FMRIB58_FA_1mm included in the FSL v5.0 software package. Resting-functional MRI: T1 template, MNI152_T1_2mm included in the FSL v5.0 software package.
Noise and artifact removal	The resting-state functional MRI data was motion corrected (Jenkinson et al., 2002) and FIX-cleaned (Salimi-Khorshidi et al., 2014) to remove physiological noise and image artefacts.
Volume censoring	Motion was corrected using MCFLIRT as implemented in FSL for the resting-state fMRI data. Diffusion MRI data was motion corrected with EDDY, also implemented in FSL. Subjects with severe head motion were not included in the preprocessed data released by the UK Biobank.

Statistical modeling & inference

Model type and settings	Multiple linear regression models, see below: Multivariate modeling and predictive analysis.
Effect(s) tested	No effects tested, because no task fMRI was performed.
Specify type of analysis:	<input type="checkbox"/> Whole brain <input checked="" type="checkbox"/> ROI-based <input type="checkbox"/> Both
Anatomical location(s)	The resting-state fMRI data were fed into an Independent Component Analysis (ICA) using the MELODIC tool (Beckmann and Smith, 2004) to identify resting-state networks present on average in the whole population. First, data was reduced to 100 dimensions using PCA and then fed into spatial ICA, from which 55 components corresponded to functional regions, and the other 45 judged to reflect physiological noise or image artifacts ("noise"). A functional component was split if it consisted of noncontiguous brain regions, yielding 81 bilateral (homotopic) regions that were further split between the hemispheres to estimate interhemispheric connectivity (see Supplementary Table 1). The 81 homotopic region pairs were all located in cortical grey matter. Cerebellar and sub-cortical components were not included in this analysis.
Statistic type for inference (See Eklund et al. 2016)	See above, clusters were obtained using spatial independent components analysis.
Correction	Statistical significance of the regression models was assessed by means of permutation testing. A null distribution was constructed for each regressor by randomly permuting the functional connectivity values (the number of permutations was set to 100,000). A p-value (two-sided) was then determined in the non-permuted model from the null distribution. Because multiple models were evaluated, we corrected for the family wise error as in (Winkler et al., 2014). Here, we generated a maximum t-statistic distribution across all homotopic region pairs and regressors (i.e., the microstructural principal components) of the permuted t-statistics. From this maximum t-statistics null-distribution a corrected p-value was estimated based for each of the non-permuted t-statistics. See Multivariate modelling and predictive analysis for more information.

Models & analysis

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Functional and/or effective connectivity
<input checked="" type="checkbox"/>	<input type="checkbox"/> Graph analysis
<input type="checkbox"/>	<input checked="" type="checkbox"/> Multivariate modeling or predictive analysis
Functional and/or effective connectivity	Functional connectivity was estimated between all pairs of homotopic regions (2x81) by means of partial correlation of the cleaned time-series using Ridge regression with a regularization factor $\rho=1$. Partial correlation aims to measure direct connectivity between two areas by first regressing out all other regions' time-series before calculating the correlation (i.e., established through inversion of the covariance matrix).
Multivariate modeling and predictive analysis	We used a multiple linear regression model to predict homotopic connectivity from a set of regressors describing the spatial pattern of microstructure along a white matter tract. The regression model was constructed for each pair of homotopic regions separately: $Y_i = X_i \beta + \epsilon_i, \quad \text{with } i=1, \dots, n$ Here Y_i ($N_{\text{subjects}} \times 1$) is a vector that contains the functional connectivity values of all subjects derived from homotopic region i (over $n = 81$ regions). To build a model using p microstructural regressors, we need to estimate a set of regression coefficients β ($p \times 1$) that describe the relative contribution from the microstructural metrics X_i ($N_{\text{subjects}} \times p$) along the white matter tract. The regressors are derived in two stages. First, the microstructural metrics were extracted from the TBSS-

voxels (white matter skeleton) corresponding to the tract of interest for every subject, yielding a matrix X^{+i} ($N_{\text{subjects}} \times N_{\text{voxels}}$). As the matrix X^{+i} is very large, a direct regression with functional connectivity is ill conditioned. We therefore perform a dimensionality reduction on X^{+i} to derive a set of regressors reflecting the primary modes of variation of a given microstructural metric across space for the cohort of subjects. A singular value decomposition (SVD) was computed from matrix X^{+i} , from which the top p components were retained, yielding matrix \tilde{X}^i ($N_{\text{subjects}} \times p$). In practice, p was set to 30 principal components, which approximately corresponded to a transition in the spectrum of singular values in terms of variance explained, above which variance explained roughly tracked noise singular vectors (Supplementary Fig. 1).

Matrices \tilde{X}^i were constructed for each of the microstructure metrics separately, yielding six single-metric linear regression models per homotopic region. In addition, a multimodal regression model was created that combined across all microstructure metrics. For the multimodal regression, all raw microstructure matrices (X^{+i}) were normalized through division by their first singular value to ensure comparable range of values. The six normalized matrices were then concatenated and an SVD was performed on the concatenated matrix to reduce back to the top 30 components.

Finally, we defined a set of confound variables of no interest (age, age², sex, age*sex, age²*sex, resting-state fMRI head motion, and head size) that could correlate with estimated microstructural measures (e.g. through artefacts such as partial volume) and thereby bias the estimated regressors. The confound variables were regressed out of the functional and microstructural data before fitting each regression model.